This research was sponsored by the
Defense Advanced Research Projects
Agency under ARPA Order No. 3690
Contract No. MDA903-79-C-0209
Monitored by Dr. J. Dexter Fletcher

Mellonics Systems Development Division
Washington Scientific Support Office
P. O. Box 1286
5265-A Port Royal Road
Springfield, Virginia 22151

Final Report

TEAM TRAINING APPLICATIONS OF VOICE PROCESSING TECHNOLOGY

Beverly A. Popelka

C. Mazie Knerr

Effective Date of Contract: 10 Jan 1979
Contract Expiration Date: 31 Mar 1980
Period of Performance: 26 Dec 78 to 29 Feb 80
Contract Period Covered by the Report: 1 Aug 79 to 31 Mar 80
Short Title of Contract: Prediction Models for Proficiency Decay

Dr. C. Mazie Knerr
Principal Investigator
321-8330

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1 REPORT NUMBER	2. GOVT ACCESSION NO. AD-A085999
3. RECIPIENT'S CATALOG NUMBER (9)	
4 TITLE (and Subtitle) (6) Team Training Applications of Voice Processing Technology,	5 TYPE OF REPORT & PERIOD COVERED Final Technical Report. 1 Aug 79-31 MAR 80
	6. PERFORMING ORG. REPORT NUMBER
7 AUTHOR(s) (10) Beverly A. Popelka [redacted] C. Mazie Knerr	8. CONTRACT OR GRANT NUMBER(s) (15) MDA903-79-C-0209 ARPA Order-3690
9. PERFORMING ORGANIZATION NAME AND ADDRESS Litton Mellonics P.O. Box 1286 Springfield, VA 22151	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS (11) Defense Advanced Research Projects Agency Cybernetics Technology Office 1400 Wilson Boulevard, Arlington, VA 22209	12. REPORT DATE 31 March 1980
	13. NUMBER OF PAGES (12) 36
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17 DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18 SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

adaptive training	skill loss	voice input
computer speech recognition	speech processing	
military training	team performance	
simulation	team performance model	
skill retention	training of teams	

20 ABSTRACT (Continue on reverse side if necessary and identify by block number)

Automated speech technology and intelligent computer assisted instruction offer unique solutions to problems of training teams in communication and coordination skills. At this point in the emergence of automated speech technology, scientists have only begun to explore its training uses. The application of automated speech technology entails adaptive training, or intelligent computer assisted instruction techniques in which the computer acts like a human tutor. This report reviews the goals and accomplishments of automated speech processing and its application to to training, especially military team training.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 68 IS OBSOLETE

409736

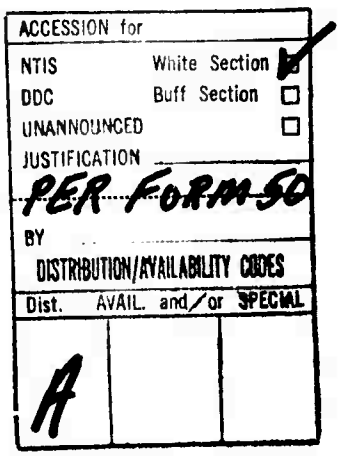
UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PREFACE

In 1976 the Defense Science Board recommended several directions for team performance research. Substantive research issues included the development of team process models; the interface between individual and team training; cost-effective team training; and improved effectiveness of simulators through application of instructional technology. Research conducted by Litton Mellonics' Washington Scientific Support Office for the Defense Advanced Research Projects Agency (DARPA) (Contract MDA903-79-C-0209) focused on team process models and the interface between individual and team skill retention. The interim report, entitled Sustaining Team Performance: A Systems Model, documented the degradation of team performance produced by team member communication requirements. The second phase of the research focused on the application of automated voice technology, employing adaptive training techniques, to improve team communication training. Given the critical role of teams in the military, application of instructional advanced technology such as voice processing deserves high priority. This final report documents the results of the second phase of the research, with Section I presenting a brief overview of the material presented in the interim report.

The authors wish to express gratitude and appreciation to the following: Dr. Dexter Fletcher at DARPA, Dr. Robert Breaux at the Naval Training Equipment Center (NTEC) and Michael Grady at Logicon for their advice and assistance in the preparation of this report.



SUMMARY

Objective

The results of the first phase of the project (Contract No. MDA 903-79-C-0209) indicated that decreases in team performance result from the effects of interactive processes such as communication and coordination. The purpose of the second phase of this research was to examine the feasibility of applying automated voice technology and adaptive training techniques to the problem of training the team in these interactive processes.

General Methods Employed

Literature on the state-of-the-art of computer speech recognition and adaptive training techniques was reviewed for the possible application of these techniques to the problem of team communication training. Literature focusing on team task analysis was also reviewed as a first step in the development of these team training systems.

Results and Conclusions

Automated speech recognition is a rapidly emerging technology that has had minimal training application. It has a major advantage of providing a medium for accurate simulation of verbal interaction. Coupled with Intelligent Computer Assisted Instruction (ICAI) techniques, this can provide a learning environment adapted to the student's individual needs. This is particularly important in the training of interactive skills, which to this point has relied on support personnel to model team members. Through the use of these combined technologies, a computer model can respond to the student's verbal commands thus minimizing uncontrolled variation in training. Improved procedures for team task analysis have made it possible to model the team, thus providing a base for adaptive training techniques. Adaptive systems based on speech understanding technology offer powerful tools for team training.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 Introduction

Empirical research demonstrates that the collective productivity of members in a team is usually less than the productivity predicted by the cumulation of individual member production. The superiority of non-interactive, individual productivity over coordinated team activity has been demonstrated for many types of tasks including problem solving (Taylor and Faust, 1952) and display monitoring (Meister, 1976; Waag and Halcomb, 1972). The decrease in performance appears to result from the effects of team interactive processes. Steiner (1966) applied the term "process loss" to the degradation of team productivity produced by interactive processes and he empirically demonstrated process loss effects (Steiner, 1972).

Communications are an example of team processes that diminish team performance. The most consistently demonstrated effect of communication on team performance is that the extent of communication is inversely related to team productivity (Johnston and Briggs, 1968; Meister, 1976; O'Brien and Owens, 1972; Steiner, 1972). Process losses appear to increase when requirements increase for communication and coordination. For example, adding members to a team degrades team performance if the addition increases the communication demands. Tasks having inherent communication demands (e.g., command and control in Army tactical operations centers and Navy combat information centers) suffer more from team process losses than do tasks without such demands. Finally, performance in emergent (unpredictable) situations is hypothesized to require more coordination and thus to have higher process losses than performance in established situations.

Team performance situations that suffer most from process losses are believed to benefit most from team training. Treatises on team training typically address the value of sequences and amounts of individual and team training (Collins, 1977; Nieva, Fleishman, and Rieck, 1978). The mediating variables seem to be the extent of communication or other interactive requirements imposed on the team. Thus, team training in communication and other interactive skills is valuable when tasks require such skills, when the work situation is emergent, and when task requirements are highly complex. According to Briggs and Johnston (1967) complexity refers to the array of stimulus inputs, control operations, and level of uncertainty in the task as a whole. They cite the operations of air traffic control and air defense systems, in contrast to controlled laboratory experiments where the superiority of individual training over team training is demonstrated.

Automated speech technology and intelligent computer assisted instruction offer unique solutions to the problems of team training in communication and coordination skills. At this point in the emergence of automated speech technology, scientists have only begun to explore its training uses. One example, a Ground Controlled Approach training system, uses automated speech technology to train an air traffic controller in the techniques of the final approach (see Section 5.1). The pilot and control tower personnel in this system are computer simulated, thus obviating the need for an instructor or support personnel to play these roles during training. By simulating all but one person in the communication chain by voice processing techniques, each student can be given individualized, adaptive instruction (Grady, Hicklin, and Porter, 1978). The system allows remedial training of one student without the time consuming involvement of support personnel. This principle can also be applied to training of military teams or crews that work together to operate or maintain weapon systems or conduct tactical missions.

The following sections describe automated voice technology and adaptive training as they apply to team training. The material on voice technology focuses on speech recognition rather than speech generation. Automated speech generation is well developed and commonly applied while speech recognition remains in developing stages despite over twenty years of research in various agencies and companies. This report focuses on recent developments: the results and recommendations of a study group convened by the Defense Advanced Research Projects Agency (DARPA) (Newell, Barnett, Forgie, Green, Klatt, Licklider, Munson, Reddy, and Woods, 1973), a special issue of the proceedings of the Institute of Electrical and Electronics Engineers (IEEE) on "Man-Machine Communication by Voice," and other books and articles. Those sources present comprehensive treatments of automated voice technology from which only the highly pertinent aspects are repeated in this report. After a minimum of basic and historical material, this report focuses on an update of applications of the technologies to team training. Similarly, it avoids repeating material presented in Litton's interim report on this same project, titled Sustaining Team Performance: A Systems Model (Knerr, Berger and Popelka, 1979).

2.0 Adaptive Training

Training systems based on the principles of artificial intelligence are variously known as: adaptive training (Knerr and Nawrocki, 1978), intelligent instructional systems (Fletcher and Zdybel, 1977), generative computer assisted instruction (Sinnott, 1976), and the all-inclusive term, intelligent computer assisted instruction (ICAI). In ICAI systems the computer acts like a human tutor. The computer is programmed with a model of the subject matter which allows it to generate information, exercises, and answers to questions by deductive inference and computation (Wolfe and Williams, 1979). Rather than tutorial or drill and practice sessions as in traditional computer assisted instruction, it provides highly individualized training without the need for specific programming for each lesson. Presentation of instructional material is adapted to the pace and instructional needs of each student. This flexibility allows each student to test hypotheses, probe for information at individual levels of difficulty and abstraction, receive instructional aids for partially completed solutions, receive reviews and critiques of completed solutions, obtain instruction generated to unique abilities and needs, and acquire a wide experience in a minimum of training time (Fletecher and Zdybel, 1977).

Fletcher and Zdybel describe the benefits of incorporating artificial intelligence algorithms into instructional media. The first benefit is the reduction of costs of instructional material preparation. Second, the training program is adapted to the specific needs of the student by automating the production of items and by branching techniques. The training objectives can be achieved with reduced training time. The last benefit is the ability to support simulations for a variety of instructional needs. The initial developmental work has been completed for instructional systems that include such diverse topics as: mathematical logic, electronic troubleshooting, instruction of BASIC, integration techniques in calculus, South American geography, Assembler language programming, and examination of student written programs in ALGOL-like language (Brown, 1977; Fletcher and Zdybel, 1977; Knerr & Nawrocki, 1978).

The Army Research Institute (ARI) applied artificial intelligence to the problem of electronic troubleshooting training (Knerr, 1979). In ARI's system, called Adaptive Computerized Training System (ACTS), the student troubleshoots an electronic circuit (making various tests, replacing a malfunctioning part, and making verification measurements). The program models the student's decision structure, compares it to the decision structure of an expert, and then adapts the instructional sequence to eliminate discrepancies between the student and expert. The program is composed of four major components: (1) the task model, a simulation of the circuit being trained; (2) the expert model, an Expected Utility (EU) decision model developed through observations of an expert's troubleshooting; (3) the student's model, an EU that predicts the student's measurement choices, developed through an on-line observation of the student's behavior; and (4) the instructional model, which determines discrepancies, modifies feedback, and modifies instructional strategies.

ACTS students select actions (e.g., take measurements, replace modules from a set of possible actions). Their actions have 96 possible outcomes. Each outcome has three properties that are combined in the Expected Utility (EU) model. The first is the conditional probability of the occurrence of that outcome given the measurement outcomes previously obtained and given that the appropriate action is selected. The second property is the subjective utility of the outcome: although subjective, it pertains to the cost (money, time) of the outcome. The third property is the gain in information that the outcome provides. The EU of the action is the sum, across all possible outcomes of that action, of the product of the three properties, stated by Knerr (1979, p. 4) as follows:

$$EU_j = \sum_{i}^{n} \alpha_{ij} U_{ij}$$

where

EU_j = the expected utility of action A_j

P_{ij} = the probability that outcome i, of a set of n outcomes will occur if action A_j is selected

U_{ij} = the utility of outcome i of action A_j

α_{ij} = the information gain resulting from the occurrence of outcome i of action A_j

The probability and information gain of outcome i of Action j (P_{ij} and α_{ij}) are constants that can be determined objectively. Values of P_{ij} and α_{ij} are entered into the model prior to the troubleshooting session, and the U_{ij} are set a common arbitrary value (usually 100). The utilities (EU_j and U_{ij}) are determined separately for each individual (expert or student) from measures of their actions during troubleshooting problems. Individuals are presented with updated probabilities of measurement outcomes (the P_{ij}) before they select a measurement.

Essential elements in adaptive training application include the task, student and ideal (expert, in ACTS) models of the behavior, and the instructional model. The instructional model requires objective automated measurement to assess student performance and to determine subsequent training trials. These adaptive training elements are necessary in systems that capitalize on voice processing technology to train voice communication-based tasks. At the same time, automated speech generation and recognition can enhance adaptive training in situations that require voice communications.

3.0 Speech Generation

Speech is a natural medium for communication between people and computers, at least from the human viewpoint. Speech has a high output rate and can be used in parallel with other media channels. Computer generated speech has been operative for years, whereas computer recognized speech has only recently become feasible.

Methods for generating speech have been very successful. Early efforts used analog recordings on a magnetic drum. These were recalled on a random access basis to generate a new phrase. Examples of this technology are the Cognitronics and Metrolab voice drums. This technology became very expensive because it did not share in the digital electronics advances in recent years. It also suffered from a segmented and choppy dialog and a very restricted vocabulary.

New electronic voice generators can synthesize phonemes and then assemble them in any sequence. This technology only requires small and inexpensive solid-state electronic systems. The advantage of this system is the unlimited vocabulary and improved voice quality through inflection, pitch and volume control. With the advent of this new technology . . . "it became literally true that it was no more difficult to cause the computer to speak than to have the computer print (although one had to learn to 'spell' all over again!)" (Grady, Hicklin & Porter, 1978, p. 101).

One commercial system that seems to be at the forefront of technology is Texas Instruments' Speak and Spell, which currently retails for approximately $50. The acoustic wave pattern is represented by 10 parameters for speech generation by the method of linear predictive coding (Robinson, 1979b). Although the general public views this educational device as an electronic toy, professionals see it as the beginning of the use of microelectronics for speech generation and recognition. The particular advantage of linear predictive coding is the minimal memory requirements for generating an impressive amount of speech.

4.0 Voice Recognition

In 1970, a study group sponsored by the Defense Advanced Research Projects Agency (DARPA) reviewed the field of automated voice processing, noted a number of considerations that had to be overcome, and made recommendations concerning the development of speech understanding systems (Newell et al., 1973). Their issues covered system characteristics (memory, processing, organization, and cost), speaker characteristics (speaker independence and gender), environmental characteristics (ambient and microphone noise), and amount of training (of the system and the user).

Some of the problems cited in the early 1970s have been solved. Voice recognition systems are currently operational in industry and, at the prototype stage, in training systems. A number of companies (Threshold Technology Interstate Electronics, Centigram Corporation, NEC America and Dialog Systems) are currently marketing voice input systems (Data Communications, 1979). A recent Associated Press article noted that a Japanese manufacturing representative (Sharp Company) predicted that "translators in which one speaks a word into the computer and gets a voice reply in the desired language will be on the market 'in the near future'" (The Sun, February 17, 1980).

Martin (1977) reviewed operational applications including quality control and inspection (television faceplate inspection, pull-ring can lid inspection, automobile assembly inspection); automated material handling (baggage, parcels, sack mail); and voice programming for numerical control (directly speaking commands to a computer rather than translating into a computer language). A voice recognition system was developed by Logicon for the Department of Transportation. This system verifies ship response to commands. The system improves safety in areas (e.g. ports) where ships must navigate in close quarters.

Successful applications in the industrial market have made voice input technology an enticing tool for training needs, especially in the field of air traffic control, where the language is very stylized. Section 5 describes some prototype training systems utilizing voice input technology.

The remainder of this section explores issues in speech understanding systems. Sections 4.1 and 4.2 present background information concerning the process of speech recognition and difference between isolated word recognition and continuous speech recognition technology. Section 4.3 discusses the advantages and disadvantages of speech input, while Sections 4.4-4.7 present the issues of system characteristics, speaker characteristics, environment and amount of training, respectively.

4.1 The Process of Speech Recognition

The first step in conversion of sound waves into meaningful communication is through the use of a transducer. Step 1 in Figure 1 represents this through the use of the human ear.* Mechanical transducers include the telephone and close speaking microphones.

The second step involves modeling of the ear through spectral analysis. It is currently felt that all acoustic information necessary for recognition can be represented by the time evolution of the power spectrum, with the phase component being relatively unimportant. Most recognition systems use some sort of spectral analysis (e.g., fast Fourier transform), but other methods may be used (i.e., autocorrelations of the amplitude, variations of the speech waveform, linear predictive codes, zero crossing statistics). These initial analyses produce parametric representations of speech in what is known as the "preprocessor." Voice input preprocessors, or sound classifiers, accomplish the front-end analysis by sampling the utterance (approximately 500 times per second) for the presence or absence of certain speech features. An example of a commercial preprocessor is discussed further in Section 4.4.

---

*This discussion is adapted from: White, George M. Speech recognition: A tutorial overview. In N. R. Dixon and T. B. Martin (Eds.), Automatic speech and speaker recognition. New York: IEEE Press, 1979 (Reprinted from Computer, 9, May, 1976).

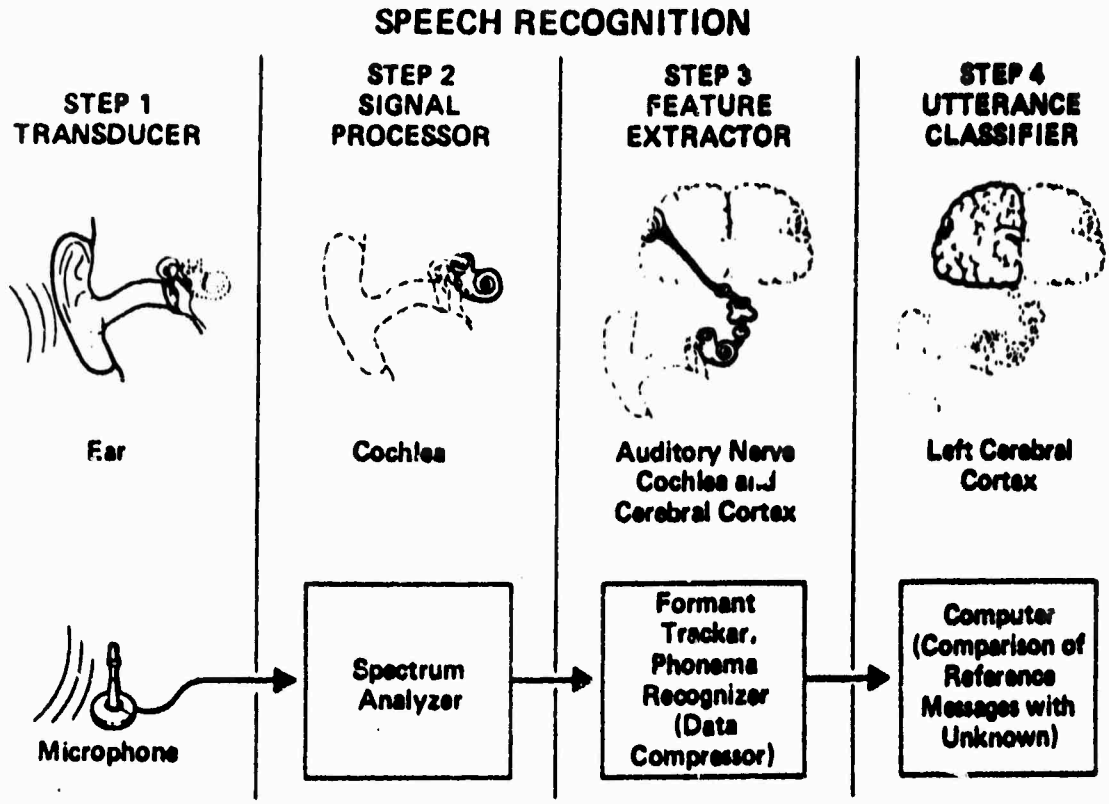


Figure 1 Overview of speech recognition showing parallels between human and machine recognition

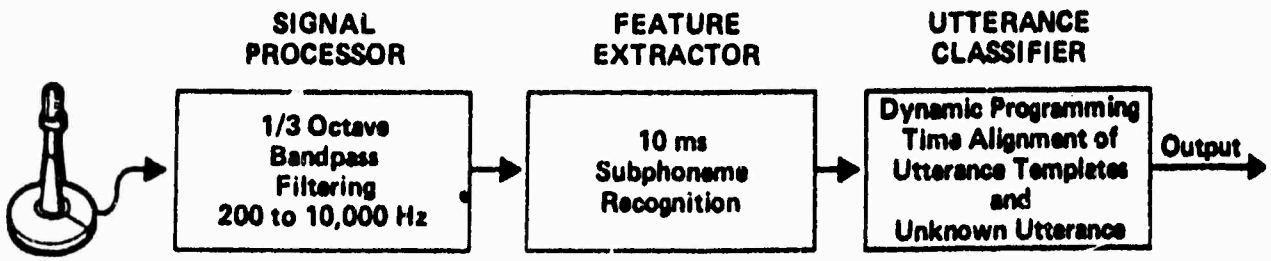


Figure 2 Flow diagram for an operational isolated utterance speech recognition, an example of the general structure

[Taken from White, George M. Speech recognition: A tutorial overview . In N. R. Dixon and T. B. Martin (Eds.), Automatic speech and speaker recognition. New York: IEEE Press, 1979, p. 90. (Reprinted from Computer, 9, May 1976.)]

Irrelevant or redundant information can then be removed through "data compression" before actual recognition occurs. Formant tracking (monitoring of the time evolution of the major peaks of the power spectrum) or recognition of subpatterns in speech are two methods of compression.

The first step in recognition (using word prototypes) is to divide the utterance into short segments (i.e., 10 ms) (Figure 2). The utterance is time-aligned (also known as time normalization) with each prototype stored in memory and the similarity between short segments in the same time interval is measured by such similarity functions as correlation, filter or geometric distance functions. The unknown utterance is then classified by the word prototype that has the best similarity score. Using entire word prototypes is inefficient and the use of subunits (e.g., phonemes) can reduce machine memory requirements.

Recognition (especially in commercial systems) is at an acoustic pattern (i.e., word template) level, while speech understanding begins at a level of recognition (i.e., word or phoneme) and utilizes higher level contextual processing of other knowledge sources. These knowledge sources include such things as semantics (meaning of sentences), syntax (grammatical struture), phonology (criteria in pronunciation), prosodics (stress and information), and context of conversation (pragmatics).

Early work assumed that higher level knowledge sources (syntax, semantics, etc.) were needed for continuous speech recognition. Processing of these knowledge sources required integration of artificial intelligence techniques into the system. Porter et al. (1977) were concerned with continuous digit entry in which these knowledge sources were irrelevant. Therefore, they did not incorporate these knowledge sources into their limited continuous recognition system. They used a simpler procedure of sequential decoding that does not involve laborious and time-consuming comparisons employed in previous systems.

4.2 Isolated Vs. Continuous Recognition

Although the isolated word technology has proven to be quite successful, the techniques developed have not solved the continuous speech problem. An additional solution proposed for continuous speech processing is the Limited Continuous Speech Recognition (LCSR) technique. These technologies differ in their assumptions and constraints as discussed briefly in the following paragraphs.

4.2.1 Isolated Word Technology

Dixon (1977) defines Isolated Word Recognition (IWR) as a process that requires the following assumptions:

- The entire utterance exists as a predefined, known image.
- It is not generally necessary to make internal decisions relative to within-image syntactic boundaries, stress and/or intonation variation.
- It is not necessary to account for variation among occurrences of an "image", other than variations in duration and amplitude.

- There is a known, finite number of utterances to be recognized, each image possessing a unique prototype in the computer memory.

Time intervals between discrete utterances is in the order of 100 ms (Martin, 1977b). Anything that is shorter in duration can be confused with stops in continuous speech, although a word such as "rapid" can be spoken with a relatively long time lapse between syllables. In operational settings, only 30-70 words per minute as an average over an 8-hour day have been achieved. These records include peak periods of 120 words per minute and lower average rates during lighter workload periods.

Although the vocabulary size is limited, entries of "words" that are to be recognized can also include not only isolated words, but phrases of short duration (a few seconds in length) as well. Early research reported very small vocabulary sizes, while Threshold and Interstate purportedly are able to handle vocabulary sizes from 600-900 words (Data Communications, August, 1979). By requiring the speaker to presegment his utterances at the word or phrase level, a pseudo-continuous speech can be achieved, using isolated word technology. This technology is congruent with stylized phrases common in air control and military radio communications ("over and out").

Isolated word recognition systems are achieving high (99%) accuracy figures in laboratory environments and have been implemented in industrial settings, while continuous speech recognition (CSR) systems are still in early prototype development.

4.2.2 Continuous Speech Recognition

Two of the major difficulties in developing continuous speech recognition have been: (1) the difficulty of determining where one word ends and another begins, and (2) acoustic characteristics of sounds and words exhibit much greater variability due to the context of the word or phrase (Reddy, 1976). True continuous speech recognition, in theory, does not require any assumptions, but Dixon (1977) lists a few constraints that have evolved:

- The 'normalcy', age, and/or sex of talker.
- The environmental conditions.
- Syntax and lexicon of 'admissible' utterances.

The level of involvement of these constraints differs with each system. HEARSAY I demonstrated live, connected speech understanding in June, 1972 (Erman, 1977). It had vocabularies of 28-76 words and operated in 7-10 times real time. Other prototype CSR systems include: DRAGON (Carnegie-Mellon); an IBM system and the Lincoln system (M.I.T. Lincoln Laboratories) (Reddy, 1976). Reddy also notes that three knowledge sources are usually required in a CSR system: phonological, lexical, and syntactical. He further defines Speech Understanding Systems (SUS) having additional requirements in that they are not task specific, do not use restricted command language, operate when an utterance is not quite grammatical or well formed and in the presence of speech-like noises (i.e., babble, cough, etc.) (Reddy, 1976). Recognition of every word is not important, but understanding of the general intent of the message is required. It is also important for the system to keep track of the context of the conversation in

order to be able to resolve any ambiguities. Prototype systems capable of context understanding include: HEARSAY II (Carnegie Mellon); SPEECHLIS (Bolt, Beranek and Newman); and VDMS (Systems Development Corporation/Stanford Research Institute) (Reddy, 1976).

4.2.3 Limited Continuous Speech Recognition (LCSR)

A potential solution for the continuous speech problem is through LCSR. This approach is particularly useful in a task specific domain with a relatively small vocabulary. It is defined as "automatically recognizing natural human speech consisting of isolated utterances which are sequences of words chosen from a small (less than 30-word vocabulary) spoken continuously, i.e., without pauses or breaks between words" (Porter, Grady and Hicklin, 1977, p. 7.). LCSR is particularly useful for recognition of numerical utterances, spoken as an unbroken sequence of digits. A need for this type of recognition capability occurs in training, where a relatively small vocabulary is used, in different combinations spoken in a natural continuous flow (e.g., spatial coordinates).

One final type of recognition system that should be noted is keyword classification. Also known as wordspotting, it is the scanning of continuous speech for the presence or recognition of a fairly stable keyword. This technology seems to have minimal applicability to training, so will not be reviewed in this report.

4.3 Advantages and Disadvantages

Verbal messages were probably the earliest form of information transfer. The verbal mode is a very comfortable and rapid means for people to convey information. Newell et al. (1973) listed times for words per second for a sample of information channels as follows:

Channel	Words/Second	
1. Reading out loud:	4	
2. Speaking spontaneously:	2.5	
3. Typing (skilled):		(5 strokes/sec.)
4. Handwriting:	.4	
5. Hand printing:	.4	
6. Telephone dialing (touch-tone)	.3	(1.5 digits/sec.)

Speed is just one advantage of a verbal mode of input. Advantages other than high data rate and ease to the human are as follows (Newell et al., 1973, p. 8):

"1. Preferred channel for spontaneous output.

2. Does not tie up hands, eyes, feet or ears.

3. Can be used while in motion.

4. Can be used easily in parallel with other channels or effectors.

5. Broadcast over short ranges (tens of feet).

6. Inexpensive and readily available terminal equipment."

In summary, speech input is favored because of the great mobility and convenience it offers. An operator of a speech input device has greater use of his hands, eyes, and other channels to monitor or transmit other information.

One of the anticipated advantages was the use of the telephone as a transducer. The telephone is not only a very inexpensive transducer with great availability, but it also foreshadows the possibility of every home owning its own computer input/output terminal. It does possess several problems, however. The telephone has a restricted bandwidth (300 to 3000 Hz) and an uneven frequency response because of the carbon microphone. Telephones contain burst noise, distortion, echo, crosstalk, frequency translation, envelope delay, clipping, and other background noises (Reddy, 1976). Even with these problems, some have had successful results with its use. However, the overwhelming method of transduction in most systems is the close speaking microphone.

Another advantage of speech technology is the application as a medium for providing simulation in training systems, particularly for tasks that require voice communication and coordination. Stylized language, grammar, and military radio techniques can be taught to trainees with a minimum amount of uncontrolled variation in performance measurement and feedback. An unbiased and consistent computer model can respond to the trainee's verbalizations and provide consistency that is not achieved with human tutors. This control of the training situation assures a standard of training that is difficult to achieve with role modeling by support personnel. The computer training system can approximate or achieve error-free learning, score performance objectively, and provide playback of the student's performance.

The disadvantage of the use of voice technology in computer training systems is student frustration and other blocks to learning if the system fails to function properly. For example, time delays frustrate the student, and failure of the system to recognize the student's verbal statements produces scoring and feedback errors. System errors can depress learning by reinforcing incorrect responses. These potential disadvantages are the reasons that system developers strive for real-time processing and high recognition reliability. High positive transfer of training from the training to the work environment is a critical measure in the evaluation of these training systems.

Disadvantages other than time delay and recognition reliability are: limited speakers allowed, extent of pretraining of the system to the user, and constrained vocabularies. Many of these obstacles are currently being overcome. The present systems still accommodate only one speaker at a time, so that their application in team training is constrained to one team member at a time.

Cost is one early factor that is rapidly decreasing, especially with the emergence of 32-bit microprocessors. Commercial systems are now available for less than $15,000. Hardware costs have been lowered in the recent past. In 1972, Hyde stated that all that was needed for laboratory investigations were: (1) a tape recorder, (2) an analog to digital converter, and (3) a general purpose computer. With the advent of better, more efficient software and microprocessors, the cost will continue to decline.

The status of system characteristics, speaker characteristics, and pre-training of the system are further explored in the following sections.

4.4 System Characteristics

Operational systems are designed for the specific task involved. As noted earlier, a tape recorder, an analog to digital converter, and a general purpose computer are needed for developmental laboratory work. Preprocessors developed by Threshold Technology are in systems developed by Logicon, National Aviation Facilities Experimental Center (NAFEC) (Connolly, 1978) and Navy training systems (Porter et al., 1977).

The preprocessors developed by Threshold sample the speech approximately 500 times per second and detect the presence of 32 speech features. These binary features are of two types: some are related to the relative energy content of specific spectral bands and others result from logical and analog operations on the short-term spectrum. This information is then relayed to a computer, where software designs complete the recognition process. They are popular because of their performance, flexibility and low cost (Grady et al., 1978). Vocabulary size and phrase length are software limited and no expensive array processors or dedicated computers are needed with these preprocessors. Grady et al. also note that the software system that they use is packaged as a FORTRAN compatible module executing under Data General Corporation's Real-time Disk Operating System (RDOS). All reference patterns can be stored on the disk and selectively retrieved in real time, thus minimizing core requirements.

A major issue is the possibility of developing continuous speech recognizers utilizing only information contained in the acoustic wave pattern (i.e. without the use of artificial intelligence techniques). A new system introduced by Nippon is one answer to this problem (Robinson, 1979a). It has a vocabulary of up to 120 words and only requires one training pass (two for digits) with a 98% accuracy in the laboratory for any combination of words that does not exceed 2.5 seconds in length. One of the major breakthroughs is the use of a technique called dynamic programming, which does not depend on a pause in the wave pattern to detect word boundaries. The response time is only a fraction of a second, but the system retails for approximately $80,000.

Tables 1 and 2 and Figure 3 represent the hardware utilized for three separate systems.

First is a research system at National Aviation Facilities Experimental Center (NAFEC) where research into application of computerized word recognition technology as an alternative to keyboard data entry was done (Connolly, 1978). Their focus has been on the adaptation and application of special technology rather than development of the technology itself.

Table 1. A System at the National Aviation Facilities Experimental Center (Connolly, 1978)

1. A basic Threshold VIP-100 model preprocessor
2. Tektronix model 4012 CRT/keyboard computer terminal (output device)
3. A 10 megabyte disc.
4. A Digital Equipment Corp. DEC-writer
5. 16K words of core memory
6. An in-house designed voice digitizer

Table 2 lists the hardware supporting a prototype trainer developed at the Naval Training Experimental Center. This trainer is currently undergoing prototype testing at the Naval Air Station in Millington, Tennessee, and is discussed in detail in Section 5.

Table 2. Navy's Ground Controlled Approach Radar Controller Training System (GCA-CTS)

1. A Threshold Technology voice input preprocessor (Threshold 500)
2. Two Data General Eclipse Computers
3. A 10M megabyte disc
4. Two CRTs
5. A Tally high speed character printer
6. A Votrax speech synthesizer
7. A Megatek MG552 display

Figure 3 displays a proposed hardware configuration for another trainer, still in developmental stages, for the Air-Intercept Controller, it is also discussed in further detail in Section 5.

In summary, the systems are simple ones to which the specialized speech processing equipment has been added.

4.5 Speaker Characteristics

A major problem for speech recognition technology is the great variability within and between speakers. (In this context, "speaker" refers to the person speaking to the system rather than a piece of equipment.) Most systems are speaker adaptive to handle the large amount of variability. Speaker adaptation entails training the system to the speech characteristics of each individual speaker using the system. In the training mode, each speaker trains the system by repeating samples of the vocabulary words. A reference set of word features is developed for each vocabulary word (or phrase) for each individual speaker. New words that are spoken into the system are compared to the reference words to achieve the best fit or a reject decision. It has also been noted that "as for operator abnormalities such as head colds, sore throats, and hoarseness, experience has shown that normally they do not affect reference patterns" (Martin, 1977a, p. 39).

Dixon (1977) noted that approximately 50 stimulus classes spoken by approximately 5 adult males were classified correctly 95% of the time. The number of speakers is inversely related to the number of recognition classes, i.e., fewer classes can be reliably classified for more speakers. Stimulus classes refer to primarily words, but short presegmented forms can also be included.

Although early studies used exclusively male speakers, Connolly (1978) reports no difference in error rates for either gender or various professions. He also notes that individual error rates ranged from less than 1% to nearly 7% regardless of gender or profession of users tested. In trying to develop a speaker independent system, Bell Laboratories is investigating the possibility of utilizing universal speech characteristics. Speech patterns of 100 men and women divided into approximately 6-12 distinct groups were analyzed (Robinson, 1979a). Within group differences were minimal. At present this system requires much computational power because the speech sample must be compared with the reference pattern for all 12 groups.

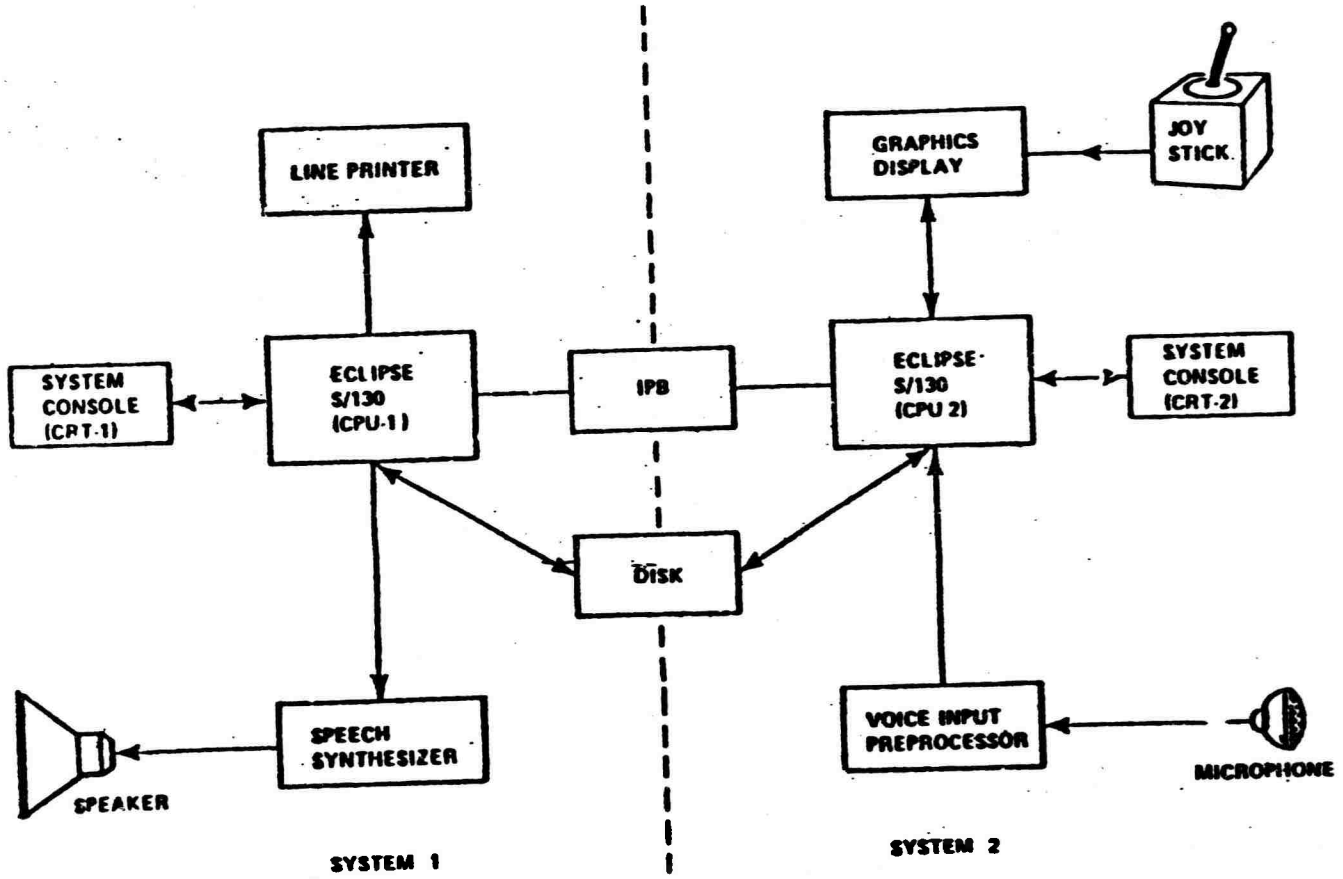


Figure 3 AIC Laboratory Hardware Configuration
(Taken from Anders et al., 1979, p. 16)

Technology and software are developing so fast that it is difficult to stay current with the state-of-the-art. Two commercial companies (Threshold and Interstate) now report 600-900 words in a structured application, while Dialog Systems has a speaker independent system that will recognize any user (Data Communications, 1979).

4.6 Environment

Environmental problems include noise and gravitational levels. Environmental noise is a major problem, but commercial systems have been developed for use in noisy environments. Special microphones reduce the amount of irrelevant ambient noise; contact microphones and close-talking, noise-cancelling microphones achieve some noise protection (Martin, 1977b). A contact microphone picks up many of the attributes of unvoiced frictional sounds and produces erroneous signals as a result of body movement. Close-talking, noise-cancelling microphones are the best means of obtaining quality speech understanding while reducing ambient noise; these can be mounted or worn on a headband.

Breath noise is a very critical problem in word recognition systems (Martin, 1977b) since it produces word boundary inaccuracies, which compound the detection problem. Separation of speech and breath noise can be achieved through pattern recognition processing. Trained speakers and motionless people produce less breath noise.

Other potential environmental problems include operator originated babble, coughs, sneezes, etc. In a limited vocabulary system, these would not be recognized as useful information and would be rejected by the system (Martin, 1977a).

Threshold Technology, Inc. has a system that is able to recognize 20 numbers with almost 99% accuracy in an airport baggage sorting operation (Martin, 1976). In this environment there exists a 90db background noise and the breath noise caused by heavy load-lifting exertion.

Other environmental concerns to the military especially are such factors as gravitational pull, voice changes due to prolonged oxygen breathing and the effects of workload stress (e.g., being shot at) Dr. Courtright noted that questions in the past were concerned with the quality of recognition in a noisy environment, but another question which must be addressed is the extent that background noise and environmental conditions affect the voice of the speaker (Naval Training Equipment Center, 1979, p. 17).

4.7 Amount of Training and the VDC System

Although some systems claim speaker independence, most systems require that each speaker train the system to recognize every word or phrase to be used. Speakers (users) repeat the words or phrases to develop a recognition template, because the system must have the voice characteristics of each speaker. This system training was an arduous task, requiring serial presentation of each vocabulary item. "The phrase was typed onto the system teletypewriter, and the student was prompted to repeat the phrase...The system collected data on the student's vocal characteristics, generated the necessary reference patterns, and moved on to the next vocabulary item" (Breaux and Grady, 1976, p. 230). System training requirements hampered user acceptance, utility, and cost (in the sense of resource time used).

The Voice Data Collection (VDC) system transformed system training into efficient use of training time. VDC was designed to enhance laboratory research and development in automated voice processing, specifically research on job vocabularies amenable to voice processing applications, examining pronunciation and radio terminology training, and analyzing vocabularies that have high-risk levels for voice processing technology. VDC acts as a peer instructor for the new student. It presents on a graphics display unit and by voice generation the vocabulary that the student must learn. It instructs the student to repeat the words and phrases after the audio and visual cues. As the student progresses, the instructor cues the system to collect voice data required for the recognition templates. Subsequently, the system collects performance data that inform the instructor whether the student's accuracy is sufficient for voice reference or whether the student needs more time to learn the vocabulary. The student learns the terminology of the task, unaware that the system is making recognition patterns of his voice. The training program proceeds (e.g., training runs on aircraft landing control) after the student and the system have both completed their pretraining. Approximately thirty minutes are required for this pretraining in the Ground Controlled Approach training system.

VDC, in essence, converts system training time into student familiarization time, an integral part of the student's program of instruction. The success of VDC has enhanced voice processing applications in Navy training, as described in the following section.

5.0 Training Systems Utilizing Speech Input

Automated speech recognition is incorporated into Air Force and Navy training systems. For example, speech technology is a small portion of the Automated Adaptive Flight Training System (AFTS) for F-14E air-to-air intercept training at 18 Air Force Bases throughout the world (Grady, personal communication). Speech technology is used more extensively in prototype systems for training two Navy jobs: ground controlled approach control and air intercept control.

Ground controlled approach and air intercept control jobs were selected for developmental systems because of the extensive voice communication requirements and the rigid and stylized vocabulary. Automation was desirable for these jobs because of the resources needed to support their training. Prior training required two support personnel for each trainee, resulting in high training costs (Breaux, 1976). The ground controlled approach and air intercept controller training systems are described in the remainder of this section.

5.1 The Ground Controlled Approach Controller Training System

The Ground Controlled Approach Controller Training System (GCA-CTS) is an experimental prototype for basic training in ground controlled aircraft approaches. The system combines automated adaptive instruction with speech recognition technology. The GCA-CTS trains Navy enlisted personnel to interpret precision approach radar (PAR) information and guide aircraft pilots in their approaches. GCA-CTS enables the instructor to serve as a training manager. It performs the routine training details such as repeating basic information, correcting errors, and maintaining training and evaluation records. The system was developed by Logicon for the Naval Training Equipment Center (NTEC) (Breaux, 1976; Goldstein et al., 1974; Grady, 1976).

The computer, employing speech recognition and speech synthesis capabilities, simulates the aircraft and control tower, roles previously played by support personnel. The isolated word recognition subsystem is based on Threshold 500 hardware, described in Section 4. Through the voice input processor, the computer "hears" what the student says (i.e., student advisories and messages to the machine-simulated pilot). The system simulates aircraft, pilot, and environmental conditions. The VDC system (Section 4.7) trains the student the basic vocabulary of GCA procedures at the same time that it collects voice data for the recognition template. Basic vocabulary training is highly suited in this situation because the system provides initial training to the students.

The student station has a graphics desplay unit for radar simulation, a cathode ray tube (CRT) to present information in text form, a keyboard for student requests, and a communication panel. The system presents training problems, or "runs" in which the student performs the behavioral sequence required on the job. For example, the student controller advises "approaching glidepath," "begin descent," and so on in response (and in control of) to the simulated landing.

The system measures the student's performance and computes a composite score. Performance measurement variables fall into six categories: safety, delivery technique, course, glidepath, landing threshold, and handoff. An example of a glidepath variable is the number of erroneous position calls (Breaux, 1976). The measures are used to provide feedback to the student and to determine the difficulty level of the subsequent run.

The adaptive training syllabus compares the student's radio transmissions to correct transmissions, measures the student's performance, and adapts the training sequence to meet the student's needs. The training program contains a variety of problems, or aircraft landing scenarios, that differ in difficulty. Difficulty is manipulated by changing pilot response, pilot variability, wind factors, and aircraft type. The system presents easier problems after a student scores poorly and harder ones after a student scores well.

Grady (1976) compared the goals of machine speech understanding stated by the Defense Advanced Research Projects Agency (DARPA) research as reported by Newell et al. in 1970 to those accomplished in the GCA-CTS. The goals set a 1976 milestone for such performance specifications as the acceptance of continuous speech in a quiet room in a few times real-time. In 1974, the GCA-CTS was ready for testing. The system accomplished most of the goals two years earlier than the 1976 milestone. GCA-CTS accepted short phrases rather than fully continuous speech, in a noisy room, in near real-time. It achieved less than 5% system response error, half the error rate goal of the DARPA research. Vocabulary limits remain a problem, with GCA-CTS permitting a highly selected vocabulary of 70 phrases rather than the 1,000 item vocabulary milestone.

In general, the GCA-CTS accomplished the automated voice processing goals for a training system and set the stage for the development of a system for air intercept control training.

5.2 Air Intercept Controller

Development of a laboratory system for Air Intercept Controller (AIC) training used the same hardware that supported the GCA-CTS experimental prototype (Anders et al., 1979). AIC phraseology is less stylized than GCA, and AIC

training is more complex, therefore AIC training increases the demands on the speech understanding subsystem. The Isolated Word Recognition (IWR) technique used in the GCA-CTS did not meet the AIC training requirements. On the other hand, continuous speech recognition is not sufficiently developed to implement. As a result, a combination of IWR and Limited Continuous Speech Recognition (LCSR) was developed and tested in the AIC training system.

The objectives of the demonstration project for the AIC system were to test the new speech understanding subsystem, evaluate training strategies, and determine the requirements, design alternatives, and costs for the simulation. One of the first tasks in the development of the AIC system was to define the training objectives and vocabulary required for AIC training. For example, the training developers wanted to train the AIC to interpret the bogey (enemy aircraft) track and ground speed symbols from the Naval Tactical Data System (NTDS) console's data readout (Anders et al., 1979). The developers designed three scenarios of increasing complexity: basic intercepts, realistic intercepts, and training environment intercepts. In the basic intercepts, the AIC student must locate his assigned (friendly) aircraft and establish radio contact. When the student detects a bogey, he guides the friendly pilot to intercept the bogey, providing information on the bogey's bearing, range, heading, and speed. The realistic intercepts add further training objectives. The AIC student must provide bogey position and velocity plus sudden changes in the bogey's heading, and must recommend new vectors in response to the changes. The AIC student must also detect and report other aircraft and respond to communications from the pilot at times in the intercept sequence that would be required on the job.

The training environment scenarios train the AIC to provide training for pilots. The AIC must be able to control two aircraft in practice intercepts in which one pilot simulates the bogey. The AIC makes radar contact with the two aircraft and establishes a lost communications procedure with the pilots. The AIC provides information to vector the aircraft for the mock intercept, then provides breakaway headings to separate the aircraft (Anders et al., 1979; Grady et al., 1979).

The developers analyzed the three scenarios to determine the transmissions required and analyzed the vocabulary in the light of technological limitations. The basic and realistic scenarios required 27 phrases (e.g., "Roger, that is your bogey tracking XXX" where XXX is a three-digit number between 001 and 360, the bearing or heading), and the training scenario required 70 phrases. Since the speech understanding technology employed was a mixed IWR/LCSR approach (rather than LCSR alone) the users had to constrain their utterances to meet the demands of isolated word recognition. For example, when they said "Bogey XXX; YY" (where XXX represents heading and YY represents range), they had to pause between XXX and YY, at the point indicated by the semicolon. Three AIC instructors tested the experimental prototype. All three stated that the constraints were not detrimental to AIC trainig, but they had difficulty initially in applying the constraints (Grady et al., 1979). A test of training transfer from the experimental to operational setting has not yet been conducted.

The developers reported problems in creating the voice data reference file and in recognition accuracy of some types of utterances (e.g., phrases ending with three digits). They recommend new algorithms for the mixed IWR/LCSR approach, improvements in LCSR techniques, and examination of other approaches (e.g., using the Nippon system for continuous speech) for AIC (Grady et al., 1979).

6.0 Training Implications

The GCA-CTS and AIC systems offer highly adaptive, individualized instruction on military tasks involving coordination and communication. These systems use the principles of individual instruction although they prepare the student to function in concert with others. The systems incorporate programs that model the ideal situation, the student's behavior, and an instructional subsystem. They provide objective measurement, feedback, and sequencing of the instructional materials based on the individual student's progress.

As Breaux (1976) notes, the training of the ground controlled approach and air intercept controllers is similar to team training. The training system in both cases teach information processing skills coordination with others. While the pairings of controllers with pilots are not teams in the strict definition of teams which requires continuity of team membership, the controller and pilot do perform coordinated, task-oriented activities to accomplish a specific objective and thus have team-like performance and training requirements.

The developmental phases required for application of voice processing in team training are similar to the phases in development of the GCA-CTS and AIC systems. However, the analysis of team tasks and development of team training materials for more complex teams or for teams operating in tactical situations are more difficult than for individual training development. The remainder of this section describes progress in the art of team task analysis and suggests some areas where voice processing may fruitfully be applied in Army training.

6.1 Team Task Analysis

Team task analysis has been explored in military contexts including Air Force team training device development (Eggemeier and Cream, 1978), Army artillery fire control training (Thurmond and Kribs, 1978), and Army training and evaluation program development (DA HQ TRADOC, undated draft). Although the analysis of team tasks is a necessary first step in training system development, the methodology is still in emerging stages.

Eggemeier and Cream (1978) expanded the traditional individual task analysis technique for use with teams. In order to select tasks to incorporate into a flight simulator for crew training, they applied three criteria: criticality, difficulty, and frequency of task performance. Over one-half of 700 tasks they identified for the training involved interaction between hardware systems, between crew members or both. In order to handle the high degree of coordination required, they separated the total aircraft mission into logical sections. Specific tasks for each team member were listed in an analysis of each segment. They described interactions among team members when performing coordinated actions, and examined actions that emerged from the combination of single tasks. Some tasks that had initially been rated low in difficulty or criticality when performed by a single crew member (or described as the task of a single crew member) were more complex and critical when performed in the mission context. Eggemeier and Cream cite several examples of complex problems that arise from coordination requirements.

Thurmond and Kribs (1978) provide additional detail concerning steps in analyzing team tasks. Using the Army computerized artillery fire control system (TACFIRE) for embedding team training, they developed sample training materials based on a team model for instructional system development. They broke

down every action into three elements: input, process, and output. Inputs are the cues or stimuli that elicit the behaviors. Processes are the behavioral steps, functions, and subtasks within the task. Task products and subsequent tasks elicited by the behaviors are the outputs. Thurmond and Kribs linked each act to other acts by either man-man or man-machine interaction. For man-man interfaces the inputs, processes, and outputs are verbal (or some representation of the verbalization) and the product of the analysis is a script of the likely or permitted statements. For many tasks these verbal transactions are highly stylized and must be conducted according to protocols, such as military status reports.

The next step in Thurmond and Kribs analysis was development of a team task flow chart and summary table. The flow chart indicates actions that emerge from the combination of tasks. As noted earlier, task combinations are particularly important in team performance since members perform tasks in concert, and individual tasks become more complex and difficult because of the combinations with other tasks being performed by other team members. The summary table listed team member involvement, type of team structure (e.g., serial or parallel team member behavior), type of interface (with other team members or with machines) and task dimensions (roles, attitudes, and communication).

The end product of this task analysis is a detailed model of the team member tasks and their interactions, analogous to the student and expert models in the ACTS program (described in Section 2).

The Army Training and Doctrine Command (TRADOC) developed a proceduralized guide for task analysis of collective (crew, unit, or team) tasks (DA HQ TRADOC, undated). The purpose of the guide is to assist Army training developers in their preparation of Army Training and Evaluation Programs (ARTEP). The procedures include analysis of the unit and analysis of the collective tasks.

The first step is to analyze battalion missions into echelon-mission-situation combinations by identifying the echelons, describing situations that each echelon may encounter, and identifying missions each echelon performs in the situations. The situations take into account the mission, threat (enemy weapons, organization, tactics, and numbers), weather and terrain conditions. Echelon-mission-situation combinations consider the situational factors for each echelon in the unit and the combined overview for the whole unit.

The next step is to analyze collective tasks by decomposing the missions until the next smaller division would be an individual task. Next, TRADOC recommends ratings of task criticality by a review board, followed by development of condition statements and standards for each critical task. The standards must be integrated to delete duplications and ensure that the standards are consistent. Next, task hierarchies are developed to show supported and supporting tasks (Figure 4). Supportability is determined by examining the unit tactics and techniques to determine how the task is performed, personnel and equipment capabilities of the unit, and the relation between collective and individual tasks. Each collective task must be supported by individual tasks.

The TRADOC procedures for collective task analysis apply to units larger, for the most part, than the situations in which automated voice processing has been applied. However, such procedures are necessary to ensure that the tasks selected for training are essential to the team's mission. The procedures described by Thurmond and Kribs appear to have the fine-grained detail applicable to team interactions suitable for voice technology applications.

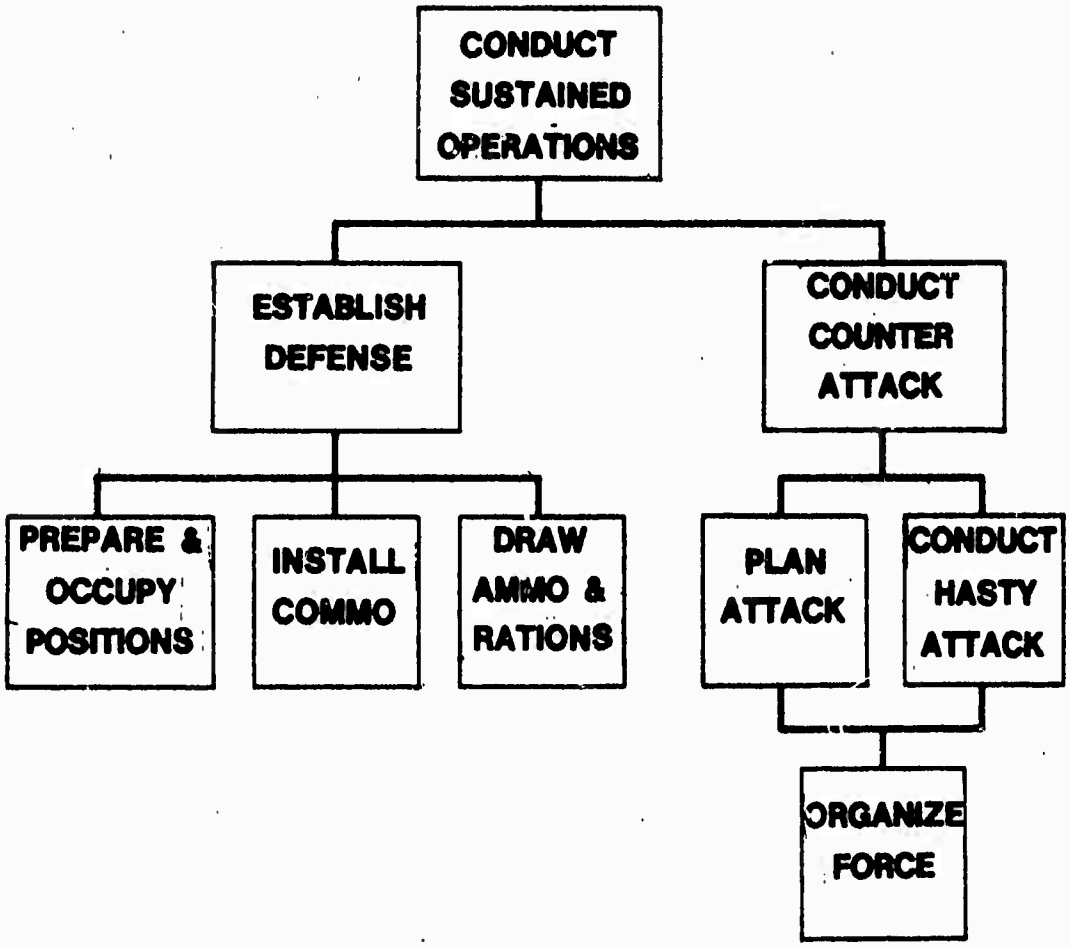


Figure 4. Task Hierarchy

Development of training materials for systems employing voice processing requires analysis of the verbalizations. The analytic process is similar to the stimulus-process-response analysis of Thurmond and Kribs but it must include all of the required and allowed utterances. Based on the analysis, the student and expert models (similar to those in the ACTS program or the GCA-CTS and AIC training systems) are developed. These models enable the instructor model to compare the student's responses to the ideal in order to measure performance objectively, control the adaptive training program, and provide feedback. The analysis of the verbalization is used to develop a lexicon containing the permitted phraseology defined for the tasks. The lexicon establishes the accepted speech patterns in the voice recognition program.

6.2 Potential Application in Army Training

Voice processing technology has been applied in Navy and Air Force training systems but not in the Army. Improvements in Army team task analysis, such as the TRADOC procedures, have set the stage for Army applications.

An example of individual tasks coordinated within an Army crew task is shown in Figure 5.

This example demonstrates how members of a team interact among themselves and with the equipment in order to perform a complex and highly critical task: firing the 155mm Howitzer, an artillery piece. There are ten members in the 155mm Howitzer (self-propelled) section:

1 = chief of section
2 = gunner
3 = assistant gunner
4-8 = cannoneers
9 = motor carriage driver
10 = section vehicle driver

Although not shown on the figure, the No. 5 cannoneer acts as the radio telephone operator and the section vehicle driver performs functions as assigned by the chief of section.

One sees in the flowchart both parallel and serial activities. Note the parallel actions that occur when the stimulus, the fire command, is received. However, also note that the projectile must be loaded, then the powder charge, then the primer must be inserted, the Assistant Gunner must call "set," then the Gunner must call "ready," and finally the chief of section can command "fire."

These are examples of man-man as well as man-machine interfaces. The accepted standard for a round to be "shot" is thirty seconds after the receipt of the quadrant elevation for the initial round and twenty seconds for subsequent rounds. The speed standard emphasizes the interdependence, sequencing, and timing that must be trained in such a team. Voice processing technology can readily be applied to training artillery crew members in the sequencing and coordination required in their jobs.

Another potential application of voice processing technology to Army training is tank crew training, where the verbal stimuli and responses in engagements are well within the state of the art of voice processing. The sequence of tank crew member tasks including verbal utterances to open a main gun engagement are:

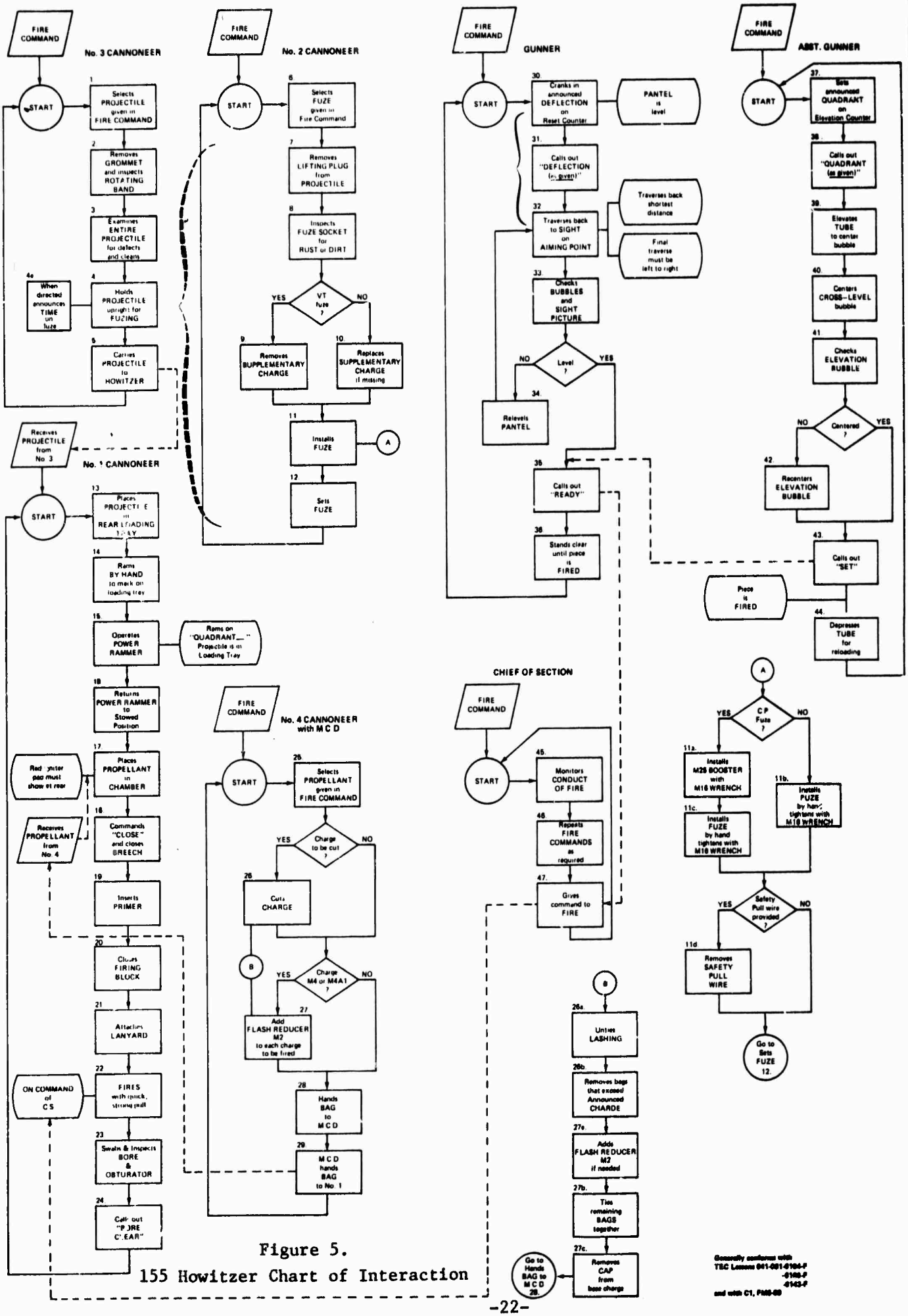


Figure 5.

155 Howitzer Chart of Interaction

CREW POSITION	ACTION
Tank commander	-commands "Driver, stop" -issues initial fire command -lays main gun approximately
Driver	-stops the tank
Gunner	-identifies target -states "Identified" -indexes ammunition -selects firing switch -makes final precise lay on target
Loader	-selects ammunition -loads gun -announces "Up"
Tank commander	-range on target -commands "Fire"
Gunner	-fires

The crew continues with subsequent firing as needed. The verbal statements, including phrases in the fire commands, could be trained to individual tank personnel to prepare them for duty in the same way that the GCA-CTS prepares aircraft controllers.

The training of coordination and communication required in Army weapon system crews is, in large part, within the capabilities of present voice processing technology. The systems developed to train Navy air intercept and ground controlled approach controllers are directly transferable to Army aircraft control. However, the Army also needs team training in situations that require advances in voice processing technology. One example is training for the tank platoon leader who must coordinate the activities of the five tanks in the platoon and must communicate with company level personnel. Other examples are training for the tactical operations center personnel as well as personnel who have other command, control, and communication functions. Training in tactical situations requires a larger scope of verbalization than do the set protocols of weapon system engagements. However, advances in hardware (e.g. the Nippon processor) and associated software may enable tactical training applications in the near future. The additional challenge in tactical training is the requirement to model the tactical behavior of the team. The model is needed as the basis for the student, ideal (expert), and instructor models to guide the adaptive training programs in a system devised for tactical training.

An advantage of this technology for training purposes is elimination of human operators. By eliminating the pseudo-pilots in the air traffic control simulation, the controllers' instructions can be translated directly into aircraft movements (Connolly, 1978). The training suffers from less human error and the training atmosphere is standardized for all students, allowing more reliable training and more accurate proficiency measurement. Breaux (1977) noted these advantages and also an increase in training per unit time compared to conventional types of training.

6.3 Other Training Applications

Connolly (1978) noted another use for speech technology in training: "with the use of reconstituted digitized spoken output (rather than synthetic speech) as an instructional and exemplary medium, such trainers would be capable of teaching not only the vocabulary and structure of the specific language, but also pronunciation and inflection" (p. 114). This is particularly pertinent for those who have a minimal grasp of the English language. Joost and Petry (1979) review a prototype system for computer-aided speech therapy. They anticipate that variations of the system will provide assistance to a wide range of speech problems, including common early childhood lisps, to training in esophageal speech. Use of computers alleviate the manpower shortage in speech pathology by relieving teachers of the time-consuming monitoring and feedback of repetitious words and phrases. Figure 6 from Joost and Petry (1979, p. 32) represents a therapeutic situation. The computer assisted interactions are shown with the dotted lines. Initial analysis and training are performed by the therapist, while the time-consuming practice sessions are computer aided. Proficiency measures that can easily be taken with this system are number of repetitions, number of acceptable responses, percentage of correct responses, or a "distance" from an ideal response.

Joost and Petry also note a few potential drawbacks. It may be difficult to elicit an acceptable model because the system is very voice-dependent, and each template must be person specific. The inflexibility of the response may not be desirable, in that reward may need to be administered to responses that do not fit the criterion, but are better than the previous response. The method of feedback (graphics, spelling, vocalization, etc.) needs to be researched, and machine prompting may be of questionable benefit.

7.0 Summary

The Defense Science Board Task Force on Training Technology (1976) noted several advances in technology, such as the laser and large-scale integration of digital circuits, that were applicable to team training in 1975. Since that time, advances in technology and reductions in cost have substantially increased the feasibility and utility of computer and other high technology resources to improve team training and team performance measurement. Among these advances are software and hardware for automated speech understanding which are particularly suited for training team members in the skills related to communication and coordination. Training systems sponsored by NTEC for the Navy jobs of ground controlled approach and air intercept control represent the forefront of speech processing technology. Speech processing has also been applied in training systems for the Air Force, but not for the Army.

Improvements in team task analysis, definition of team performance variables, and other aspects of instructional systems development for team training have paralleled the technological advances. Procedures such as those outlined by the TRADOC guide for collective task analysis link the team or crew tasks to the unit's mission. Thurmond and Kribs (1978) provide more detailed steps, analyzing the stimuli, processes, and responses within the team task. Eggemeier and Cream (1978) also provide detail concerning the analysis of individual behaviors in team tasks, and they describe actions and difficulties that emerge from combinations of single tasks. Used in conjunction with adaptive training techniques, team task analysis (particularly analysis of the communication and coordination demands) and speech understanding systems offer powerful tools for the improvement of team training.

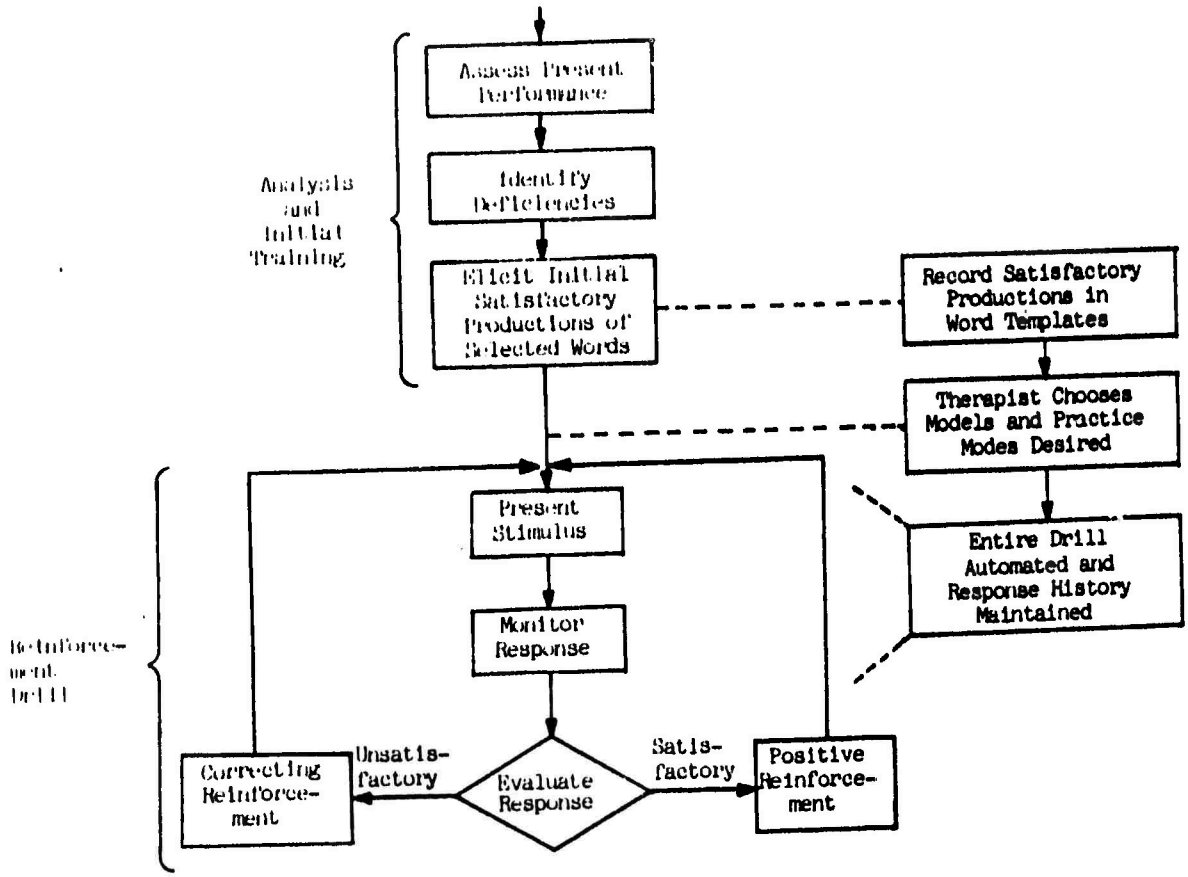


Figure 6 Therapy Procedure
(Taken from Joost and Petry, 1979, p.32)

ANNEX A

BIBLIOGRAPHY

Anders, R. M., Grady, M. W., Nowell, L. H., and Overton, M. A. A laboratory system for Air Intercept Controller training (Technical Report NAVTRAEQUIPCEN 78-C-0053-1). Orlando, FL: Naval Training Equipment Center, January 1979. (NTIS No. AD A069060)

Breaux, R. Training characteristics of the Automated Adaptive Ground Controlled Approach Radar Controller Training System (GCA-CTS) (Technical Report NAVTRAEQUIPCEN TN-52). Orlando, FL: Naval Training Equipment Center, July 1976.

Breaux, R. Laboratory demonstration of computer speech recognition in training. In Proceedings of the NTEC/Industry Conference, November 1977, pp. 169-172.

Breaux, R. and Grady, M. W. The voice data collection program: A generalized research tool for studies in speech recognition. In Proceedings of the NTEC/Industry Conference, November 1976, pp. 229-234.

Briggs, G. E. and Johnston, W. A. Team training (Technical Report 1327-4) Orlando, FL: Naval Training Device Center, June 1967.

Brown, J. S. Uses of artificial intelligence and advanced computer technology in education. In R. J. Seidel and M. L. Rubin (Eds.), Computer and communications: Implications for education. New York: Academic Press, 1977.

Collins, J. J. A study of the potential contributions of small group behavior research to team training technology development (Final Report ONR-00014-76-C-1076, NR-179-34). Alexandria, VA: Essex Corporation, August 1977.

Connolly, D. W. Speech recognition: A review of studies at NAFEC. In Proceedings: Technology in Air Traffic Control Training and Simulation. Warrenton, VA: Society for Applied Training Technology, February 1978, II, 110-115.

Defense Science Board. Crew/group/unit training. In Report of the Task Force on Training Technology. Washington, D.C.: Office of the Director of Defense Research and Engineering, Department of Defense, February 1976.

Department of the Army, Headquarters, United States Army Training and Doctrine Command. TRADOC Pamphlet 310-8 (Draft), Collective Front-End Analysis (CFEA) for Development of Army Training and Evaluation Program (ARTEP). Fort Monroe, VA: United States Army Training and Doctrine Command, undated draft.

Dixon, N. R. Automatic recognition of contiruous speech: Status and possibilities for an operational system. In R. J. Seidel and M. L. Rubin (Eds.), Computers and communications: Implications for education. New York: Academic Press, 1977.

Eggemeier, F. T. and Cream, B. W. Some considerations in development of team training devices. In Psychological Fidelity in Simulated Work Environments, Donald E. Erwin (Ed.), Proceedings of a Symposium, American Psychological Association. Toronto, Canada: August 1978.

Erman, L. D. Speech understandings systems: Hearsay and some prognostications. In R. J. Seidel and M. L. Rubin (Eds.), Computers and communications: Implications for education. New York: Academic Press Inc., 1977.

Fletcher, J. D. and Zdybel, F. Intelligent instructional systems in military training. Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977. (ED 143-371)

Goldstein, D., Norman, D. A., Charles, J. P., Feuge, R. L., Grady, M. W., Barkovic, M. H. Ears for automated instruction systems: Why try? In NTEC/Industry Conference Proceedings: Training Economy Through Simulators, November 1974, pp. 253-267.

Grady, M. W. Advanced speech technology -- The natural man/machine interface. In Second International Conference on System, Man, and Cybernetics, IEEE Proceedings, 1976, pp. 1-4.

Grady, M. W., Hicklin, M. B., and Porter, J. E. Practical applications of interactive voice technologies -- Some accomplishments and prospects. In Proceedings: Technology in Air Traffic Control Training and Simulation. Warrenton, VA: Society for Applied Learning Technology, February 1978, II, pp. 100-109.

Grady, M. W., Porter, J. E., Satzer, Jr., W. J., and Sprouse, B. D. Speech understanding in air intercept controller training system design (Technical Report NAVTRAEQUIPCEN 78-C-0044-1). Orlando, FL: Naval Training Equipment Center, January 1979. (NTIS No. ADA068612)

Hyde, S. R. Automatic speech recognition: A critical survey and disucussion of the literature. E. E. David, Jr. and P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw Hill, 1972, pp. 339-438.

Johnston, W. A. and Briggs, G. E. Team performance as a function of team arrangement and workload. Journal of Applied Psychology, 1968, 52 (2), 89-94.

Joost, M. G. and Petry, F. E. Design factors in the use of isolated word recognition for speech therapy. In Proceedings of the Human Factors Society -- 23rd Annual Meeting, 1979, pp. 30-34.

Knerr, B. W. Adaptive computerized training system (ACTS): Relationships between utility similarity and strategy similarity (Research Memorandum 79-5). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, April 1979.

Knerr, B. and Nawrocki, L. H. Development and evaluation of an adaptive computerized training system (R & D utilization report 78-1). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, September 1978.

Knerr, C. M., Berger, D. C., and Popelka, B. A. Sustaining team performance: A systems model. Interim Report. DARPA Contract No. MDA903-79-C-0209. Springfield, VA: Litton Mellonics, March 31, 1980.

Martin, T. B. Practical applications of voice input to machines. Proc. IEEE, 1976, 64, 487-501.

Martin, T. B. A practical voice input system. In R. J. Seidel and M. L. Rubin (Eds.), Computers and communications: Implications for education. New York: Academic Press, 1977.

Martin, T. B. One way to talk to computers. IEEE Spectrum, May 1977, pp. 35-39.

Meister, D. Behavioral foundations of system development. New York: John Wiley, 1976.

Naval Training Equipment Center. Voice technology for system application. Minutes of a Sub Technical Advisory Group (SUB TAG). San Antonio, TX: March 6-8, 1979.

Newell, A., Barnett, J., Forgie, J. W., Green, C., Klatt, D., Licklider, J. C. R., Munson, J., Reddy, D. R., and Woods, W. A. Speech understanding systems. New York: American Elsevier Publishing Company, Inc., 1973.

Nieva, V. F., Fleishman, E. A., and Rieck, A. Team dimensions: Their identity, their measurement and their relationships. Advanced Research Resources Organization. Princeton, NJ: Response Analysis Corp., November 20, 1978.

O'Brien, G. E. and Owens, A. G. Effects of organizational structure on correlations between member abilities and group productivity. In B. M. Bass and S. D. Deep (Eds.), Studies in organizational psychology. Boston, MA: Allyn and Bacon, Inc., 1972.

Porter, J. E., Grady, M. W., Hicklin, M. B., and Lowe, L. R. Use of computer speech understanding in training: A preliminary investigation of a limited continuous speech recognition capability (Technical Report NAVTRAEQUIPCEN 74-C-0048-2). San Diego, CA: Logicon, Inc., 1977. (NTIS No. ADA049680)

Reddy, D. Speech recognition by machine: A review. Proc. IEEE, 1976, 64, 501-531.

Remote data entry recognizes a strong new voice. Data Communication, August 1979, pp. 15-18.

Robinson, A. L. More people are talking to computers as speech recognition enters the real world. Science, February 16, 1979, pp. 634-638.

Robinson, A. L. Speech is another microelectronics conquest. Science, February 16, 1979, p. 635.

Sinnott, L. T. Generative computer-assisted instruction and artificial intelligence (ARPA Order No. 2651). Princeton, NJ: Educational Testing Service, October 1976. (NTIS No. ADA040963)

Steiner, I. D. Models for inferring relationships between group size and potential group productivity. Behavioral Science, 1966, 11, 273-283.

Steiner, I. D. Group process and productivity. New York: Academic Press, 1972.

Taylor, D. W. and Faust, W. L. Twenty questions: Efficiency in problem solving as a function of size of group. Journal of Experimental Psychology, 1952, 44, 360-368.

Thurmond, P. and Kribs, H. D. Computerized collective training for teams (ARI Final Report TR-78-A1). Alexandria, VA: Sensors, Data, Decisions, Inc., February 1978.

U.S. made first computerized translator but Japanese are capturing the market. The Sun, February 17, 1980.

Waag, W. L. and Halcomb, C. G. Team size and decision rule in the performance of simulator monitoring teams. Human Factors, 1972, 14 (4), 309-314.

White, M. Speech recognition: A tutorial overview. In N. R. Dixon and T. B. Martin (eds.) Automatic speech and speaker recognition. New York: IEEE Press, 1979. (Reprinted from Computer, 9, May 1976.)

Wolfe, J. H. and Williams, M. D. Prospects for low cost advanced computer-based training: A forecast (NPRDC Special Report 79-16) San Diego, CA: Navy Personnel Research and Development Center, April 1979.